35P

School of Electrical Engineering
Electronics Systems Research Laboratory
Purdue University
Lafayette, Indiana

Memorandum Report 64-1 May 29, 1964

THRESHOLD SELECTION AND DETECTOR PERFORMANCE WITH UNKNOWN NOISE COVARIANCE

by

W. D. Gregg

Under the Direction of
Professor J. C. Hancock

SUPPORTED BY THE
NATIONAL AERONAUTICAL AND SPACE ADMINISTRATION,
GRANT - NsG-553

THRESHOLD SELECTION AND DETECTOR
PERFORMANCE WITH UNKNOWN NOISE COVARIANCE

W. D. GREGG

28842

Summary: Particular results of Keehn [3] are applied to the problem of coherent detection of white, bandlimited, stationary, gaussian signals in additive white, bandlimited, stationary, gaussian noise with unknown covariance. The results indicate that the critical term in the aposteriori probability expression, arising from Bayes' theorem, is the Wishart density, $W(2WT, n_1, \Phi_1)$, and is maximized when the norm, $\| \underline{Z}-\underline{S}_i \|$ is minimized. The results are extended to include an illustration of threshold selection and an evaluation of detector performance. Threshold selection based upon estimation leads to detector performance that is a random variable. The probability law for detector performance and the expected performance and variance of performance are developed. Demonstration of the convergence of the expected performance and variance of performance to the specified performance and zero respectively is given. An approximation of the probability law associated with a hypothesis testing variable is used to obtain closed form expressions. An investigation of error in the approximation is included.

Author

I. INTRODUCTION

Signal detection in digital data systems generally consists of comparing some computed quantity with one or more thresholds in order to establish the message symbol transmitted. Since message symbols are encoded into channel symbols for modulation and transmission, the object of the detection system is to process the received waveforms and to make the decision. Processing is necessary because the channel characteristics alter the transmitted signals.

Data processing of received waveforms is carried out with the intent of extremizing some criterion of optimality for some available signal-to-noise ratio, SNR, and for some specified data rate (binary digits or code word blocks per unit time). The criterion of optimality has been generally the maximization of the aposteriori (sometimes called forward) probability obtained with Bayes' theorem. This criterion is exactly identical to the information theory concept of maximizing the mutual or transinformation relative to a source (transmitter) and a sink (receiver). Thus Bayes' theorem specifies the data-processing structure of the detector.

II. FORMULATION OF THE DETECTION PROBLEM IN TRUE DECISION SPACE

The majority of the problems under consideration are centered about such models as illustrated in Fig. 1.

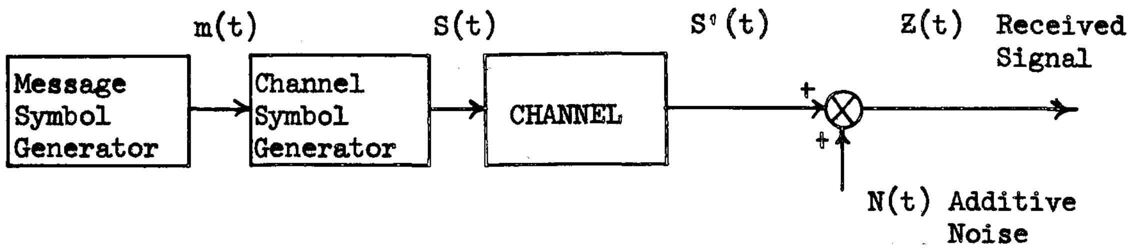

Fig. 1

The channel characteristics may consist of multipath propagation with various degrees of delay for each path, multiplicative disturbances which may vary with path delay, phase shift, and additive noise. The simplest situation consists of additive noise and is illustrated in Fig. 2, which

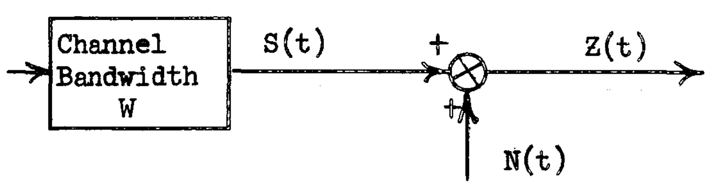

Fig. 2

represents the situation under consideration.

Recognizing that a time limited function $s(t)$ <u>cannot</u> be represented exactly by a finite combination of orthonormal basis functions from a set that is complete and closed in a harmonic extension of $s(t)$, it is still possible to approximate the norm of $s(t)$ by a finite number of terms.

With a bandlimited assumption and invoking the sampling theorem, we see that for a baud of duration T sec., one obtains 2WT samples. Assuming s(t) and n(t) are sample functions from independent, stationary, gaussian, random processes, bandlimited to W cps; they may be represented by vectors $\underline{S}$ and $\underline{N}$ in a vector space of dimension 2WT. Since $\underline{Z} = \underline{S} + \underline{N}$, Bayes' theorem is

$$P\left\{ \underline{S}_i \mid \underline{Z} \right\} = \frac{P\left\{\underline{S}_i\right\} P\left\{\underline{Z} \mid \underline{S}_i\right\}}{P\left\{ \underline{Z} \right\}} \tag{1}$$

and i=1, 2 for the binary channel.

Then for identically distributed samples,

$$P\left\{ \underline{Z} \mid \underline{S}_i \right\} = [2\pi]^{\frac{-2WT}{2}} \left| \Lambda_n \right|^{-\frac{1}{2}} e^{-\frac{1}{2}[\underline{Z}-\underline{S}_i]' \Lambda_n^{-1} [\underline{Z}-\underline{S}_i]} \prod_{m=1}^{2WT} dz_m \tag{2}$$

$$P\{\underline{Z}\} = [P\{\underline{S}_1\}\ P(\underline{Z} \mid \underline{S}_1) + P\{\underline{S}_2\}\ P(\underline{Z} \mid \underline{S}_2)] \prod_{m=1}^{2WT} dz_m \tag{3}$$

Thus

$$P\left\{ S_i \mid \underline{Z} \right\} = \frac{P\{\underline{S}_i\}[2\pi]^{-\frac{2WT}{2}} \left| \Lambda_n \right|^{-\frac{1}{2}} e^{-\frac{1}{2}[\underline{Z}-\underline{S}_i]' \Lambda_n^{-1} [\underline{Z}-\underline{S}_i]}}{[P\{\underline{S}_1\}\ P(\underline{Z} \mid \underline{S}_1) + P\{\underline{S}_2\}\ P(\underline{Z} \mid \underline{S}_2)]} \tag{4}$$

For purposes of comparing $P\{\underline{S}_1 \mid \underline{Z}\}$ with $P\{\underline{S}_2 \mid \underline{Z}\}$ for a given received baud, $\underline{Z}$, the only term of importance is the exponent weighed by appropriate values of $P\{\underline{S}_1\}$, $P\{\underline{S}_2\}$ which are usually assumed equal for the equally probable binary case. Thus the ith detector branch is to minimize the quantity $[\underline{Z} - \underline{S}_i]' \Lambda_n^{-1} [\underline{Z} - \underline{S}_i]$. With an additional assumption of white gaussian processes,

$$R_{\underline{Z}}(\tau) = K_o W \frac{\sin(2\pi W\tau)}{2\pi W\tau} \tag{5}$$

which has zeros at points $\tau = {}^n/2W$, n=1, 2, ... Hence sampling at a rate 1/2W sec. yields uncorrelated and independent samples.

The quadratic form to be minimizing thus becomes

$$\frac{[\underline{z} - \underline{S}_i]' [\underline{z} - \underline{S}_i]}{\sigma_n^2} = \frac{\|\underline{z} - \underline{S}_i\|}{\sigma_n^2} \qquad (6\text{-A})$$

The same result of uncoupling the quadratic terms could have been achieved by "pre-whitening" requiring an orthogonal transformation on $\underline{z}$. For a binary channel, the detector computes $\|\underline{z} - \underline{S}_1\|$ and $\|\underline{z} - \underline{S}_2\|$ and selects the smallest. However,

$$\|\underline{z} - \underline{S}_i\| = \sum_{m=1}^{2WT} z_m^2 - 2z_m s_{im} + s_{im}^2 \qquad (6\text{-B})$$

and it is necessary to have phase coherency to carry out the computation, e.g., the receiver must have time reference knowledge relative to the received waveform amplitude. Thus minimizing the norm, (6-B), consists of maximizing the inner product

$$\sum_{m=1}^{2WT} z_m \, s_{im} = 2W \int_0^T z(t)\, s_i(t)\, dt \qquad i = 1, 2 \qquad (7)$$

The detector structure for this simplified case, with known covariance, is then a correlation computer. For the binary channel, the Bayes' detector structure is illustrated in Figs. 3 and 4 for continuous and discrete data processing respectively.

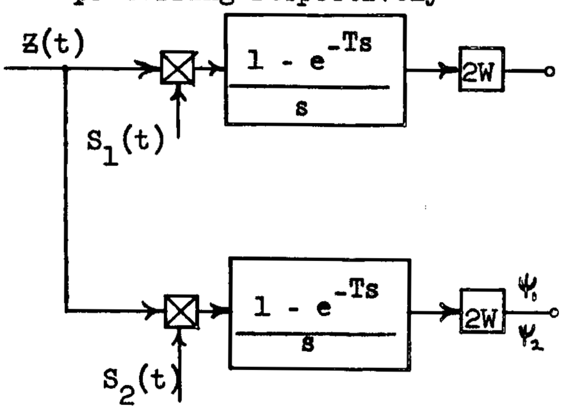

Fig. 3

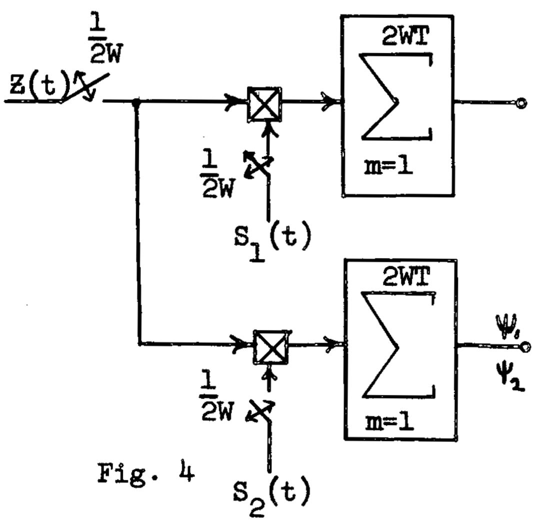

Fig. 4

Assuming uncorrelated signals, $S_1(t) = 0$; $s_2(t) = s(t)$; $0 \leq t \leq T$, and in view of the assumption on $s(t)$ and $n(t)$, the detector computed quantities will be random variables with means and variances for signal absent and signal present respectively,

$$E[\psi_1] = 0 \tag{8}$$

$$E[\psi_2] = 2WT\,\sigma_s^2 = 2WT(S_oW) \text{ watts} \tag{9}$$

and

$$\sigma_{\psi_1}^2 = 2WT\,\sigma_n^2\,\sigma_s^2 \tag{10}$$

$$\sigma_{\psi_2}^2 = 2WT\,\sigma_s^2\,[2\sigma_s^2 + \sigma_n^2] \tag{11}$$

which are derived in Appendix I.

Normalizing with respect to 2WT one obtains Y_1 and Y_2. Note that since Y_1 is a linear combination of received data whereas Y_2 is the sum of squares, we have $P(Y_1) = N(0,\sigma_n^2\sigma_s^2)$ and $P(Y_2)$ a gamma density which is asymptotically normal for large 2WT, $N(\sigma_s^2, \sigma_s^2(2\sigma_s^2 + \sigma_n^2))$.

The <u>true</u> decision space is a vector space of dimension 2WT and the computed quantities Y_1, and Y_2 represents <u>distances</u> in the vector space from $E[Y_1]$ and $E[Y_2]$. These distances are random variables with the aforementioned probability law and can be mapped onto the line Y_i, $i = 1,2$. A sketch of the p.d f. of Y_1 relative to the p.d.f. of Y_2 is illustrated in Fig. 5.

If the parameters σ_s^2 and σ_n^2 <u>are known</u>, then a threshold, Λ_o, can be selected to yield any desired false alarm probability, α_o, or false dismissal probability β_o, or for equal α and β, according to

$$\alpha_o = \int_{\Lambda_o}^{\infty} P(Y_1)\,dY_1 \implies \beta = \int_{o}^{\Lambda_o} P(Y_2)\,dY_2 \tag{12}$$

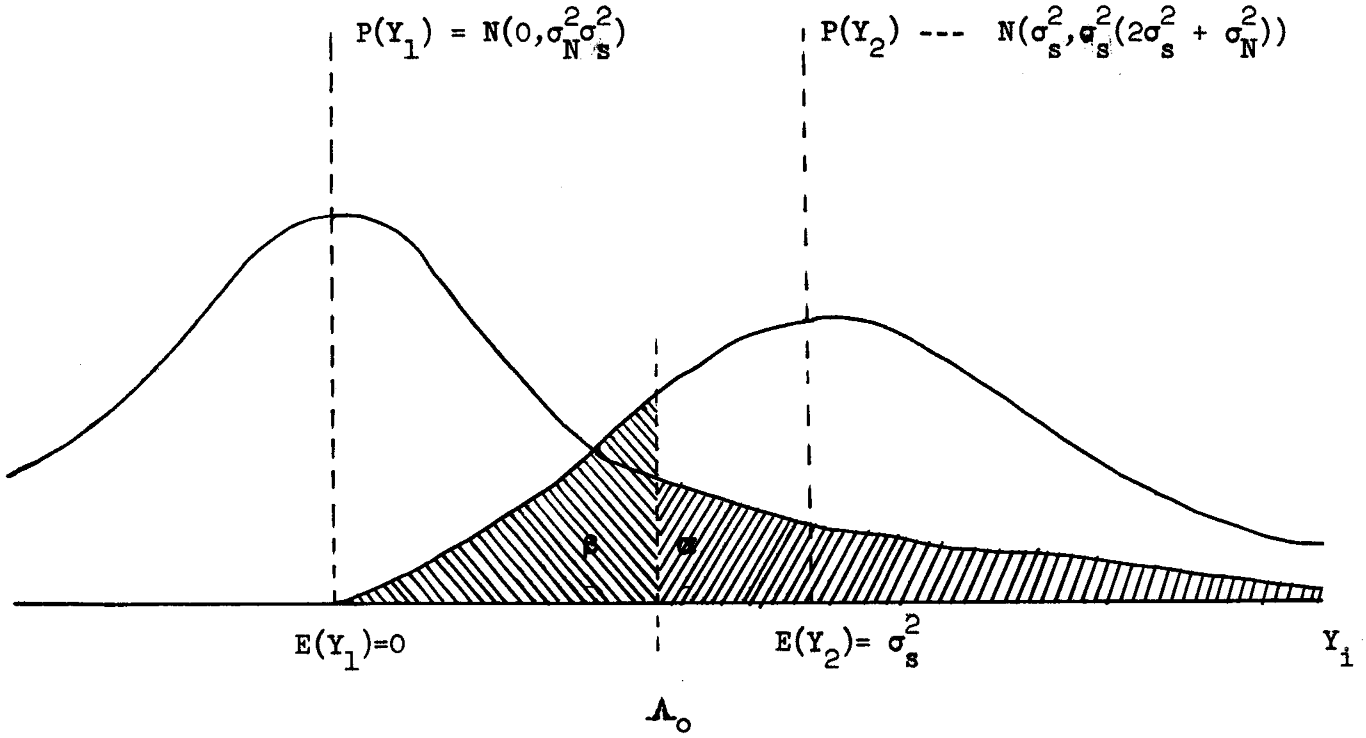

Fig. 5

or

$$\beta_o = \int_0^{\Lambda_o} P(Y_2)dY_2 \Longrightarrow \alpha = \int_{\Lambda_o}^{\infty} P(Y_1)dY_1 \tag{13}$$

or Λ_o such that

$$\alpha_o = \int_{\Lambda_o}^{\infty} P(Y_1)dY_1 = \int_0^{\Lambda_o} P(Y_2)dY_2 = \beta_o \tag{14}$$

for some specified α_o or β_o where $0 < \alpha_o < \frac{1}{2}$, $0 < \beta_o < \frac{1}{2}$.

It is reasonable to assume knowledge of s(t) at the receiver such that σ_s^2 is known. However, the covariance matrix of the additive channel noise is not known, hence the true decision space is unknown and any threshold selection based upon estimation leads to random detector performance.

It is the object of the following section to establish the form of the optimum Bayes' detector structure for the case of unknown noise using supervisory data transmissions over a bauds.

III. STRUCTURE OF THE OPTIMUM BAYES' RECEIVER WITH UNKNOWN CHANNEL NOISE COVARIANCE MATRIX

In situations where additive noise is stationary or "slowly varying", estimates of the noise covariance, Λ, can be made. Slowly varying can be interpreted as that situation, such that when ample samples of a process are taken to insure convergence in probability of the computed statistic, no appreciable difference is observed in the computed statistic.

The estimation properties of a parameter $\theta = \Lambda^{-1}$, depends upon the sufficiency, consistency, and efficiency of the estimator $\hat{\theta}$. The estimator $\hat{\theta}$ is usually some transformation or combination of samples from a population with parameter θ, and converges to θ in the limit. The property of convergence is utilized to reduce the uncertainty of θ by reducing the variance of $\hat{\theta}$. Thus, given a sequence of samples $\{\underline{z}_j\} = \underline{z}_1\ \underline{z}_2 \ldots \underline{z}_n$ where

$$\underline{z}_j = \begin{bmatrix} z_{1j} \\ \vdots \\ z_{kj} \end{bmatrix}$$

A model can be established where the decision on the j^{th} baud must be obtained with no knowledge of θ except the estimate $\hat{\theta}$ obtained from the current baud and j-1 previous bauds (see Fig. 6).

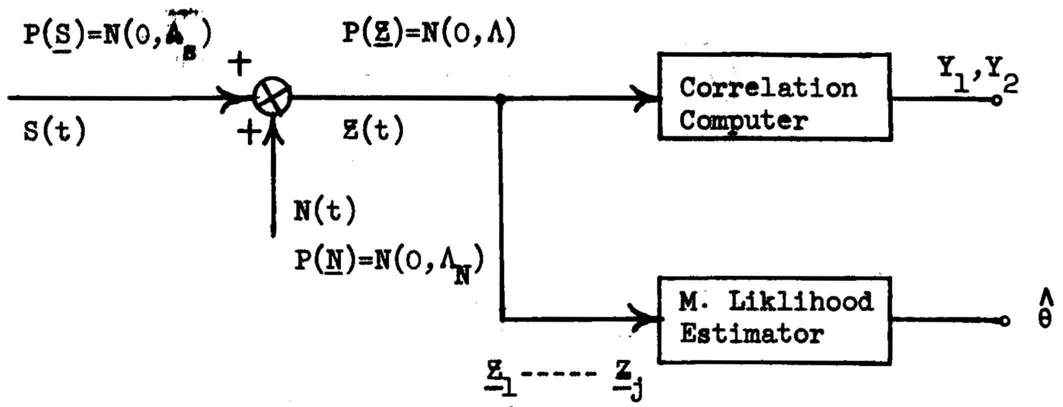

Fig. 6

This model represents the situation where a supervisory period of transmission is provided in order to establish θ to within certain confidence limits. The supervisory transmission implies <u>classification</u> of the received signal $\underline{\mathcal{Z}}_j$ at the receiver on a per baud basis.

Bayes' theorem for this situation is

$$P\{\underline{S}_i | \underline{Z};\underline{\mathcal{Z}}\} = \frac{P\{\underline{S}_i\}\ P(\underline{Z};\ \underline{\mathcal{Z}} | \underline{S}_i)}{[P\{\underline{S}_1\}\ \ P(\underline{Z},\underline{\mathcal{Z}} | \underline{S}_1) + P\{\ \underline{S}_2\}\ P(\underline{Z},\underline{\mathcal{Z}} | \underline{S}_2)]} \qquad (15)$$

and is maximized when $P(\underline{Z},\underline{\mathcal{Z}} | S_i)$ is maximized, thus maximizing the aposteriori probability of $\underline{S}_i$ being transmitted given the j^{th} baud $\underline{\mathcal{Z}}$ and $j - 1$ previous bauds reflected in $\underline{Z} = \underline{\mathcal{Z}}_1, \ \ldots, \ \underline{Z}_j$

However

$$P(\underline{Z},\underline{\mathcal{Z}} | \underline{S}_i) = P(\underline{\mathcal{Z}} | \underline{Z};\underline{S}_i)\ P(\underline{Z} | \underline{S}_i) \qquad (16)$$

and

$$P(\underline{\mathcal{Z}} | \underline{Z};\underline{S}_i) = \int P(\underline{\mathcal{Z}} | \underline{Z};\theta,\underline{S}_i)\ d\theta \qquad (17)$$

Since $P(\underline{Z} | \underline{S}_i)$ does not affect the decision, the Bayes' receiver must compute only (17).

The first term in the integrand is

$P(\underline{\mathcal{Z}} | \underline{Z};\theta;\underline{S}_i)$, which is the same as $P(\underline{\mathcal{Z}} | \theta;\underline{S}_i)$, where θ is the inverse of the covariance matrix of the process $\mathcal{Z}(t)$ when the i^{th} signal is given.

For the Gaussian assumption,

$$P(\underline{Z} \mid \theta;\underline{S}_i) = (2\pi)^{-\frac{k}{2}} \mid \theta \mid^{1/2} e^{-\frac{1}{2}[\underline{Z}-\underline{S}_i]'\theta[\underline{Z}-\underline{S}_i]} \tag{18}$$

The second term in the integrand is

$$P(\theta \mid \underline{Z};\underline{S}_i) = \frac{P(\theta)\, P(\underline{Z} \mid \theta;\underline{S}_i)}{\int P(\underline{Z} \mid \theta;\underline{S}_i)\, P(\theta)d\theta} \tag{19}$$

The probability law on θ, $P(\theta)$, if found via the Wishart law [1,2]; and using a slight modification of Keehn's result [3] to reflect apriori knowledge is

$$P(\theta) = C_{k,n_o} \left| \frac{n_o}{2} \Phi_o \right|^{\frac{n_o-1}{2}} \left| \theta \right|^{\frac{n_o-k-2}{2}} e^{-\frac{1}{2}\mathrm{TR} n_o \Phi_o \theta} \tag{20-A}$$

where

$$C_{k,n_o} = \frac{1}{\pi^{\frac{k(k-1)}{4}} \prod_{q=1}^{k} \Gamma\left(\frac{n_o-q}{2}\right)} \tag{20-B}$$

is the required normalizing term developed by Wishart to insure that $P(\theta)$ satisfies the axioms required of a p.d.f.

(See Appendix II for a complete interpretation and derivation of the results in this section).

The term Φ_o in 20-A can be interpreted as an estimate of θ based upon n_o samples from a population containing θ or simply an initial guess weighed by some constant n_o.

Using (18)

$$P(\underline{Z} \mid \theta;\underline{S}_i) = \prod_{j=1}^{n} P(\underline{Z}_j \mid \theta;\underline{S}_i) = (2\pi)^{-\frac{nk}{2}} \left| \theta \right|^{\frac{n}{2}} e^{-\frac{1}{2}\mathrm{TR} n\theta < [\underline{Z}-\underline{S}_i][\underline{Z}-\underline{S}_i]' >} \tag{21}$$

for the $\underline{z}_j$ identically distributed.

Neglecting the denominator of (19) since it only serves to normalize $P(\theta|\underline{Z};\underline{S}_i)$ to satisfy $\int..\int P(\theta|\underline{Z};\underline{S}_i)\, d\theta_{11}\ldots d\theta_{kk} = 1$, and using (20-A) and (21), and combing terms

$$P(\theta|\underline{Z};\underline{S}_i) \simeq P(\theta)P(\underline{Z}|\theta;\underline{S}_i)$$

$$= \left\{C_{k,n_o}\,(2\pi)^{-\frac{nk}{2}}\left|\frac{n_o}{2}\Phi_c\right|^{\frac{n_o-1}{2}}|\theta|^{\frac{n_o+n-k-2}{2}}\right.$$

$$\left. e^{-\frac{1}{2}\,TR\frac{(n_o+n)\theta(n_o\Phi_o+n<[\underline{Z}-\underline{S}_i][\underline{Z}-\underline{S}_i]'>)}{(n_o+n)}}\right\} \qquad (22\text{-A})$$

Thus $P(\theta|\underline{Z};\underline{S}_i)$ is a Wishart density, $W(k, n_1, \Phi_1)$ with parameters

$$n_1 = n_o + n$$

$$\Phi_1 = \frac{n_o\Phi_o + n<[\underline{Z}-\underline{S}_i][\underline{Z}-\underline{S}_i]'>}{n_o+n} \qquad (22\text{-B})$$

Substituting n_1, Φ_1 into (22-A) and modifying C_{k,n_o} to absorb the effects of the substituion and retain normality yields

$$P(\theta|\underline{Z};\underline{S}_i) = C_{k,n_1}\left|\frac{n_1}{2}\Phi_1\right|^{\frac{n_1-1}{2}}|\theta|^{\frac{n_1-k-2}{2}}\,e^{-\frac{1}{2}TRn_1\Phi_1\theta} \qquad (22\text{-C})$$

Using (18) and (22-C), the integral of (17) is

$$P(\underline{z}|\theta;\underline{S}_i)P(\theta|\underline{Z};\underline{S}_i) = (2\pi)^{-\frac{k}{2}}\,C_{k,n_1}\left|\frac{n_1}{2}\Phi_1\right|^{\frac{n_1-1}{2}}\cdot$$

$$e^{-\frac{1}{2}TR[(\underline{z}-\underline{S}_i)(\underline{z}-\underline{S}_i)'+n_1\Phi_1]\theta}|\theta|^{\frac{n_1-k-1}{2}} \qquad (23)$$

In order to evaluate an integral of the form of (17), Keehn uses the identity

$$\int\ldots\int|\theta|^{\frac{n_1-k-1}{2}}\,e^{-TRA\theta}d\theta_{11}\ldots d\theta_{kk} = \frac{1}{C_{k,n_1}|A|^{\frac{n_1}{2}}} \qquad (24)$$

which yields the original quantity to be maximized

$$P(\underline{Z}|\underline{Z};\underline{S}_i) = \frac{(2\pi)^{-\frac{k}{2}}\left|\frac{n_1}{2}\Phi_1\right|^{\frac{n_1-1}{2}}}{\left|\frac{1}{2}([\underline{Z}-\underline{S}_i][\underline{Z}-\underline{S}_i]' + n_1\Phi_1)\right|^{\frac{n_1}{2}}} \tag{25}$$

However

$$\left|\frac{1}{2}([\underline{Z}-\underline{S}_i][\underline{Z}-\underline{S}_i]' + n_1\Phi_1)\right|^{\frac{+n_1}{2}} = \left|\frac{n_1}{2}\Phi_1\right|^{\frac{n_1}{2}}\left(1 + \frac{1}{n_1}\mathrm{TR}(\underline{Z}-\underline{S}_i)(\underline{Z}-\underline{S}_i)'\Phi_1^{-1}\right)^{\frac{+n_1}{2}} \tag{26}$$

With (26), and factoring and normalizing (25) [see Appendix II starting with equation II-19] (25) becomes

$$P(\underline{Z}|\underline{Z};\underline{S}_i) = \frac{\Gamma\left(\frac{n_1}{2}\right)(2\pi)^{-\frac{k}{2}}\left|\Phi_1^{-1}\right|^{-\frac{1}{2}}}{\Gamma\left(\frac{n_1-k}{2}\right)\left(\frac{n_1}{2}\right)^{\frac{1}{2}}}\left(1 + \frac{1}{n_1}\mathrm{TR}\,(\underline{Z}-\underline{S}_i)(\underline{Z}-\underline{S}_i)'\Phi_1^{-1}\right)^{\frac{-n_1}{2}} \tag{27}$$

which for the band-limited assumption,

$K = 2WT$ (baud sample size)

n = number of supervisory bauds currently received; range of j in the sequence $\{\underline{Z}_j\}$

Also,

$$\mathrm{TR}(\underline{Z}-\underline{S}_i)(\underline{Z}-\underline{S}_i)'\Phi_1^{-1} = (\underline{Z}-\underline{S}_i)'\Phi_1^{-1}(\underline{Z}-\underline{S}_i) \tag{28}$$

Consequently the aposteriori probability is <u>maximized</u> when (28) is <u>minimized</u>. Recalling (22-B), and the assumptions of Section II, if the estimate Φ_o is obtained from the population, or simply assumed to be diagonal if Φ_o is a guess, Φ_1^{-1} will also be diagonal and then (28) will be minimum when $||\underline{Z}-\underline{S}_i||$ is minimized as in section II. Hence for white noise, the Wishart density <u>also</u> leads to the correlation computer receiver structure as the optimum detector structure.

Keehn's[3] work essentially develops the form of the Bayes' receiver structure for unknown covariance matrix yielding a Wishart Density in the aposteriori probability expression. This form, assuming signal classification on a per baud basis at the receiver, leads to the same form of detector structure in the i^{th} receiver branch as for the case of known covariance matrix with one exception. The exception is that the covariance matrix Φ_1, which is a weighed average of Φ_o and $<[\underline{Z}-\underline{S}_i][\underline{Z}-\underline{S}_i]'>$, is used in place of the true covariance Λ. Φ_o represents apriori knowledge about Λ and the term $<[\underline{Z}-\underline{S}_i][\underline{Z}-\underline{S}_i]'>$ represents corrections to Φ_o based upon n samples from n k-variate bauds. Φ_o is either a guess or an apriori estimate of Λ based upon n_o samples.

In summary, this section has thus developed the Wishart density term of the aposteriori probability expression for the case of <u>known</u> signal covariance and <u>unknown</u> noise covariance. It is the object of the following section to illustrate a technique of threshold selection and to establish the type of detector performance that might be expected.

IV. THRESHOLD SELECTION AND DETECTOR PERFORMANCE WITH UNKNOWN NOISE COVARIANCE

For the conditions and assumptions established in Section II, the p.d.f.'s associated with the hypothesis testing variables, Y_1 and Y_2, were observed to be $N(0,\sigma_n^2\sigma_s^2)$ and gamma, becoming asymptotically normal $N(\sigma_s^2,\sigma_s^2(2\sigma_s^2+\sigma_n^2))$, respectively. Equations (12), (13), and (14) were given to illustrate a means of selecting the threshold, Λ_o, for partitioning the true decision space when signal and noise parameters are known.

For supervisory data transmission, it is appropriate to indicate the threshold by Λ_{on}, since it is to be updated at the end of each baud. Since the noise parameter Λ_N is not known, it is necessary to estimate this parameter utilizing all classified data available in $\underline{Z}$, and any other apriori knowledge available in the form of Φ_o.

Recalling (22-B) and noting that

$$< [\underline{z}-\underline{S}_i][\underline{z}-\underline{S}_i]' > = \frac{1}{n}\sum_{j=1}^{n} [\underline{z}-\underline{S}_i]_j[\underline{z}-\underline{S}_i]_j' = \Lambda \qquad (29\text{-A})$$

is a weighted sum of n symmetric, kxk matrices which add up to one matrix with the ℓ^{th} diagonal elements given by

$$\lambda_{\ell\ell} = \frac{1}{n}\sum_{j=1}^{n}\left\{(z_\ell - s_{i\ell})^2\right\}_j \qquad \ell=1,\dots,k \qquad (29\text{-B})$$

For the assumptions of white Gaussian signal and noise, the variation about zero of the non-diagonal terms, which are estimates of the covariances of Λ_N, will be of no interest.

Sampling at a rate of 1/2w will yield samples spaced in time consistent with the zeros of the autocorrelation function and hence independent. For the identically distributed samples (29-A) will reduce to a scalar matrix and assuming the same characteristics for the apriori knowledge Φ_o yields,

$$\Lambda_N = \frac{n_o \Lambda_o + n\,\Lambda}{n_o + n} = \left(\frac{n_o\,\hat{\sigma}^2_{N_o} + n\,\hat{\sigma}_N^2}{n_o + n}\right) I \qquad (29\text{-C})$$

The combined estimate for the noise parameter $\sigma_N^{\,2}$ is then

$$\hat{\sigma}^2_{1N} = \frac{n_o\,\hat{\sigma}^2_{N_o} + n\,\hat{\sigma}^2_N}{n_o + n} \qquad (29\text{-D})$$

where $\hat{\sigma}^2_{N_o}$ is an apriori estimate from a population containing σ^2_N, based upon n_o samples and $\hat{\sigma}^2_N$ is an estimate based upon n classified samples.

If an unbiased estimator is used for both estimates, then

$$E[\hat{\sigma}^2_{1N}] = \frac{n_o\,E[\hat{\sigma}^2_{N_o}] + n\,E[\hat{\sigma}^2_N]}{n_o + n} = \frac{n_o\,\sigma^2_N + n\,\sigma^2_N}{n_o + n} = \sigma^2_N \qquad (29\text{-E})$$

and σ^2_{1N} is unbiased for σ^2_N. If $\sigma^2_{N_o}$ is a guess with weighting n_o it will appear as a bias but will be negligible for large n.

Selection of a threshold for some specified false alarm probability, say α_o, requires the solution of

$$\alpha_o = \int_{\Lambda_{o_n}}^{\infty} \frac{1}{\sqrt{2\pi}\,M\,\sigma_s}\, e^{-\frac{Y_1^{\,2}}{2M^2\sigma_s^{\,2}}}\, dY_1 \qquad (30\text{-A})$$

where M is the greatest upper baud, g.u.b., with confidence level p

$0 < p < 1$

written

$$\frac{(n_o+n)\ \hat{\sigma}_N^{\ 2}}{x^2_{u_{n_1}}} \leq \sigma_N^{\ 2} \leq \frac{(n_o+n)\ \hat{\sigma}_N^{\ 2}}{x^2_{L_{n_1}}} = M \tag{30-B}$$

if $\hat{\sigma}_{N_o}^{\ 2}$ is an estimate based upon n_o apriori samples. The chi-square, x^2, quantities arise from $\int^{x^2_{L_{n_1}}} f(x^2)dx^2 = \int_{x^2_{u_{n_1}}}^{\infty} f(x^2)dx^2 = \frac{1-p}{2}$, $o < p < 1$

$$\text{where } f(x^2) = \frac{e^{-\frac{x^2}{2}}\ (x^2)^{\frac{n_1-1}{2}-1}}{\Gamma\left(\frac{n_1-1}{2}\right)} = G\left(x^2; \frac{1}{2}, \frac{n_1-1}{2}\right)\ 0 < x^2 < \infty \tag{31}$$

(32)

If $\hat{\sigma}_{N_o}^{\ 2}$ is a guess no consistent quantatative confidence can be assigned.

Using maximum likelihood-unbiased estimation, the solution of

$$\frac{\partial}{\partial \sigma_{N_\ell}} \text{ Log } P(\underline{Z}_1 \cdots \underline{Z}_n; \Lambda_N) = 0 \tag{33-A}$$

$$\text{yields } \hat{\sigma}^2_{N_\ell} = \frac{1}{n-1} \sum_{j=1}^{n} [z_{\ell_j} - \overline{z_\ell}]^2 \text{ where } \overline{z_\ell} = \frac{1}{n} \sum_{j=1}^{n} z_{\ell_j} \tag{33-B}$$

which is sufficient for σ_N since

$$\begin{aligned} P(\underline{Z}_1, \ldots, \underline{Z}_n; \sigma_N^2) &= \frac{1}{(\sqrt{2\pi}\ \sigma_N)^{nk}}\ e^{-\frac{k}{2\sigma_N^2} \sum_{j=1}^{n} \underline{Z}_j^{\ 2}} \\ &= \frac{1}{(\sqrt{2\pi}\ \sigma_N)^{nk}}\ e^{-\frac{k}{2\sigma_N^2} [\sum_{j=1}^{n} (z_j - \overline{z})^2 + n\ \overline{z}^2]} \\ &= \frac{1}{(\sqrt{2\pi\sigma_N})^{nk}}\ e^{-\frac{k}{2\sigma_N^2} [(n-1)\sigma_N^2 + n\ \mu_N^{\ 2}]} \\ &= g_1(\hat{\sigma}_N^{\ 2}, \hat{\mu}_N^{\ 2}; \sigma_N^{\ 2}, 0)\ h(\underline{Z}) \end{aligned} \tag{34}$$

where $h(\underline{Z})$ is independent of $\sigma_N^{\;2}$ and degenerates to 1 in this case. Dropping the ℓ subscript as a result of the identical distribution assumption, the estimate $\hat{\sigma}_{1N}^{\;2}$ is a random variable with probability density function

$$f(\hat{\sigma}_{1N}^{\;2}) = G(\hat{\sigma}_{1N}^{\;2}\,;\ \frac{n_1-1}{2\sigma_N^{\;2}},\ \ \frac{n_1-1}{2}\ ) \qquad 0 < \hat{\sigma}_N^{\;2} < \infty \tag{35}$$

Thus M will be a random variable with a gamma probability density and the initial step is to solve (30-A) for Λ_{o_n}. Rewriting (30-A) as

$$\alpha_o = \frac{1}{2}\left[1 - \frac{2}{\sqrt{\pi}} \int_o^{\frac{\Lambda_{on}}{\sqrt{2M\sigma_s}}} e^{-w^2}\,dw\right] = \frac{1}{2}\left[1 - \text{erf}\left(\frac{\Lambda_{o_n}}{\sqrt{2M\sigma_s}}\right)\right] \tag{36}$$

the difficulty is immediately observed in that the error function over a finite interval is an incomplete gamma function and only an approximate solution will be possible.

Proceeding directly, there are two possible series expressions for (36). One is an asymptotic expansion about ∞ for large Λ_{o_n}, hence small α_o, which is found to be

$$\alpha_o = \frac{1}{2\sqrt{\pi}}\, e^{-\frac{\Lambda_{o_n}}{2M^2\,\sigma_s^2}} \sum_{\ell=o}^{\infty} \frac{(-1)^{\ell}\left(\frac{1}{2}\right)_{\ell}}{\left[\frac{\Lambda_{o_n}}{\sqrt{2M\sigma_s}}\right]^{2\ell+1}} \tag{37}$$

with the generalized factorial term $\left(\frac{1}{2}\right)_{\ell} = \prod_{k=1}^{\ell} \left(\ \frac{1}{2}\ \right) + k - 1),\ \ \ell \geq 1$; $\left(\frac{1}{2}\right)_{\ell} = 1$, $\ell = o$

and which can be truncated with an error less than the <u>last</u> term used.

The other series is obtained by a direct term by term integration of the convergent series for the integral of (36) to obtain

$$\alpha_o = \frac{1}{2} - \frac{1}{\sqrt{\pi}} \sum_{\ell=o}^{\infty} \frac{\left(\frac{\Lambda_{on}}{\sqrt{2M\sigma_s}}\right)^{(2\ell+1)} (-1)^{\ell}}{(2\ell+1)\,!} \tag{38}$$

A point solution for Λ_{on} is desired as opposed to an interval solution. In order to obtain this from (37) or (38), a single term truncation would be necessary which obviously is not satisfactory. Thus no further useful progress using the direct approach with (37) or (38) seems possible. However the theory of approximation suggests a change of integrand in (30-A) with a suitable approximation in situations such as this. Thus by approximating $P(Y_1)$ with the Laplace density function

$$P(Y_1) = \frac{1}{2a} e^{-\frac{|Y_1 - \mu|}{a}} \qquad - \infty < Y_1 < \infty \quad (39)$$

$$\mu = o, \; a = \frac{M\sigma_s}{\sqrt{2}}$$

it is possible to obtain a point solution for Λ_{on} from (30-A) as

$$\Lambda_{on} = \frac{\sqrt{n_1}}{\sqrt{2}} \quad \frac{\sigma_s(-\ln 2\alpha_o)\,\hat{\sigma}_{1N}}{x_{L_{n1}}} > 0 \qquad (40)$$

which is positive since all terms are positive and α_o is such that $0 < \alpha_o < \frac{1}{2}$. Furthermore, end conditions are also consistent in that, for some specified $\alpha_o \to 0$; $\Lambda_{on} \to +\infty$ and for $\alpha_o \to \frac{1}{2}$, $\Lambda_{on} \to 0$ as required (see Fig. 5).

Thus since Λ_{on} is a linear function of $\hat{\sigma}_N$, taking on new, updated, values at the end of each received baud, it will be a random variable following (35) since combinations of gamma variables reproduce gamma variables. Consequently, it follows that the actual probability of false alarm, α, will also be a random variable.

At this point it is necessary to disgress and outline the remaining objectives. First, it is necessary to establish, through a series of transformations, a probability density function of a variable as characteristic of α as possible; then to establish the convergence properties of the mean and variance of this variable.

The actual value of α at any decision time is given by

$$\alpha = \int_{\Lambda_{o_n}}^{\infty} \frac{1}{\sqrt{2}\,\sigma_N \sigma_s}\, e^{-\frac{Y_1 \sqrt{2}}{\sigma_N \sigma_s}}\, dY_1 \tag{41}$$

which yields

$$(\mathrm{Ln}\,\frac{1}{2\alpha})^2 = h_o \frac{n_1 \hat{\sigma}_{1N}^2}{\sigma_N\, x^2_{L_{n_1}}} = h_o \frac{S_1^2}{S_2^2}\;; \quad h_o = (\mathrm{Ln}\,\frac{1}{2\alpha_o})^2 \tag{42}$$

Equation (42) is written as a linear combination of a ratio of second degree random variables which suggests the change of variable

$$\Delta = \frac{S_1^{\,2}}{S_2^{\,2}} \quad \text{where}$$

$$P(S_1^{\,2}) = G\left(\frac{n_1-1}{2\sigma_N^{\,2}}\;;\; \frac{n_1-1}{2},\, S_1^{\,2}\right) \qquad 0 < S_1^{\,2} < \infty \tag{43-A}$$

and

$$P(S_2^{\,2}) = G\left(\frac{n_1-1}{2\sigma_N^{\,2}}\;;\; \frac{n_1-1}{2},\, S_2^{\,2}\right) \qquad 0 < S_2^{\,2} < \infty$$

Using an identity [4], the density

$$f(\Delta) = \int_{\Delta = \frac{S_1^2}{S_2^2}} G\left(\frac{n_1-1}{2\sigma_N^{\,2}}\;;\; \frac{n_1-1}{2},\, S_1^{\,2}\right) G\left(\frac{n_1-1}{2\sigma_N^{2}}\;;\; \frac{n_1-1}{2};\, S_2^{\,2}\right) dS_1^{\,2} dS_2^{\,2} =$$

$$\frac{\Gamma(n_1-1)\, \Delta^{\frac{n_1-1}{2} - 1}}{\left\{\Gamma\left(\frac{n_1-1}{2}\right)\right\}^2 (1+\Delta)^{n_1-1}} \tag{43-B}$$

is obtained.

Now let $h = (\mathrm{Ln}\,\frac{1}{2\alpha})^2 = h_o\,\Delta$ (44)

Then

$$P(h) = P(\Delta)\,|J|\,\Big|_{\Delta = \frac{h}{h_o}} \qquad |J| = \frac{1}{h_o} \tag{45}$$

leads to

$$P(h) = P([\ln \frac{1}{2\alpha}]^2) = \left[\frac{1}{\ln \frac{1}{2\alpha_o}}\right]^{\frac{n_1-1}{2}} \frac{\Gamma(n_1-1)\, h^{\frac{n_1-1}{2}-1}}{[\Gamma(\frac{n_1-1}{2})]^2 (1+\frac{h}{h_o})^{n_1-1}} \qquad 0 < h < \infty \tag{46}$$

From (46) and (47), using tables [5],

$$E[h] = [\ln \frac{1}{2\alpha_o}]^2 \frac{\Gamma(\frac{n_1+1}{2})\, \Gamma(\frac{n_1-1}{2}-1)}{[\Gamma(\frac{n_1-1}{2})]^2} \tag{48-A}$$

and

$$E[h^2] = [\ln \frac{1}{2\alpha_o}]^4 \frac{\Gamma(\frac{n_1+3}{2})\, \Gamma(\frac{n_1-5}{2})}{[\Gamma(\frac{n_1-1}{2})]^2} \tag{48-B}$$

thus

$$\sigma_h^2 = \frac{[\ln \frac{1}{2\alpha_o}]^4}{[\Gamma(\frac{n_1-1}{2})]} [\Gamma^2(\frac{n_1-1}{2})\, \Gamma(\frac{n_1+3}{2})\, \Gamma(\frac{n_1-5}{2}) - \Gamma^2(\frac{n_1-3}{2})\, \Gamma^2(\frac{n_1+1}{2})] \tag{49}$$

For large values of n_1,

$$\frac{\Gamma(\frac{n_1-1}{2}-1)\, \Gamma(\frac{n_1+1}{2})}{\Gamma^2(\frac{n_1-1}{2})} \to 1 \tag{50}$$

and

$$\Gamma^2(\frac{n_1-1}{2})\, \Gamma(\frac{n_1+3}{2})\, \Gamma(\frac{n_1-5}{2}) - \Gamma^2(\frac{n_1-3}{2})\, \Gamma^2(\frac{n_1+1}{2}) \to 0 \tag{51}$$

Thus in the limit, as $n_1 \to \infty$

$$E[h] = E[(\ln \frac{1}{2\alpha})^2] \to (\ln \frac{1}{2\alpha_o})^2 \tag{52}$$

and

$$\sigma_h^2 \to 0 \tag{53}$$

Curves of $E[h] / (\ln \frac{1}{2\alpha_o})^2$ and $\sigma_h^2 / (\ln 1/2\alpha_o)^4$ as a function of n_1 are obtained from (48-A) and (49) and illustrated on Figs. 7 and 8.

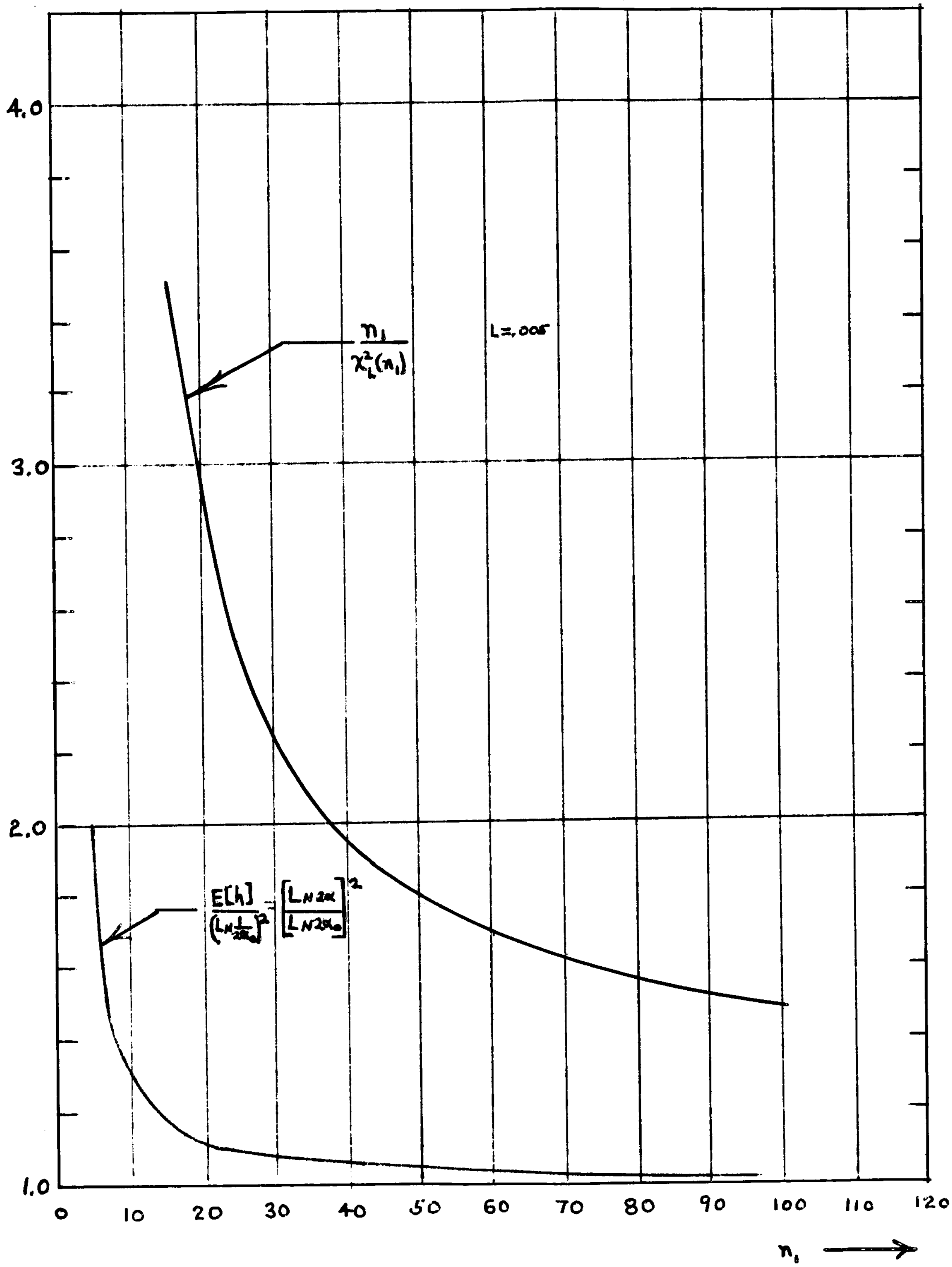
4.0
3.0
2.0
1.0
0
10
20
30
40
50
60
70
80
90
100
110
120
$\frac{n_1}{\chi^2_L(n_1)}$
L=.005
$\frac{E[h]}{(L_N\frac{1}{2\alpha_0})^2} = \left[\frac{L_N 2\alpha}{L_N 2\alpha_0}\right]^2$
n_1

Fig. 7

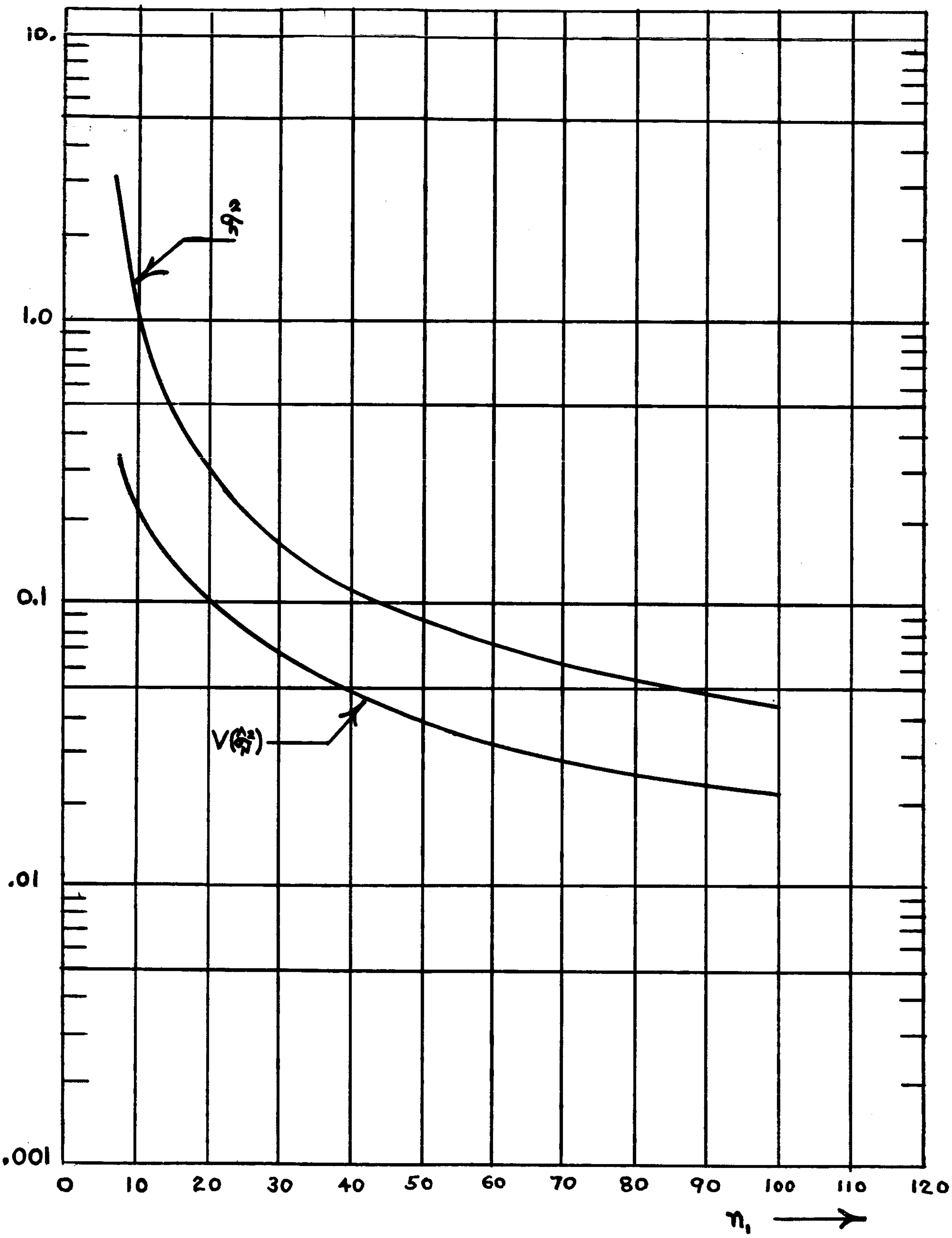

Fig. 8

The required receiver-detector structure for up dating Λ_{on} is illustrated on Fig. 9.

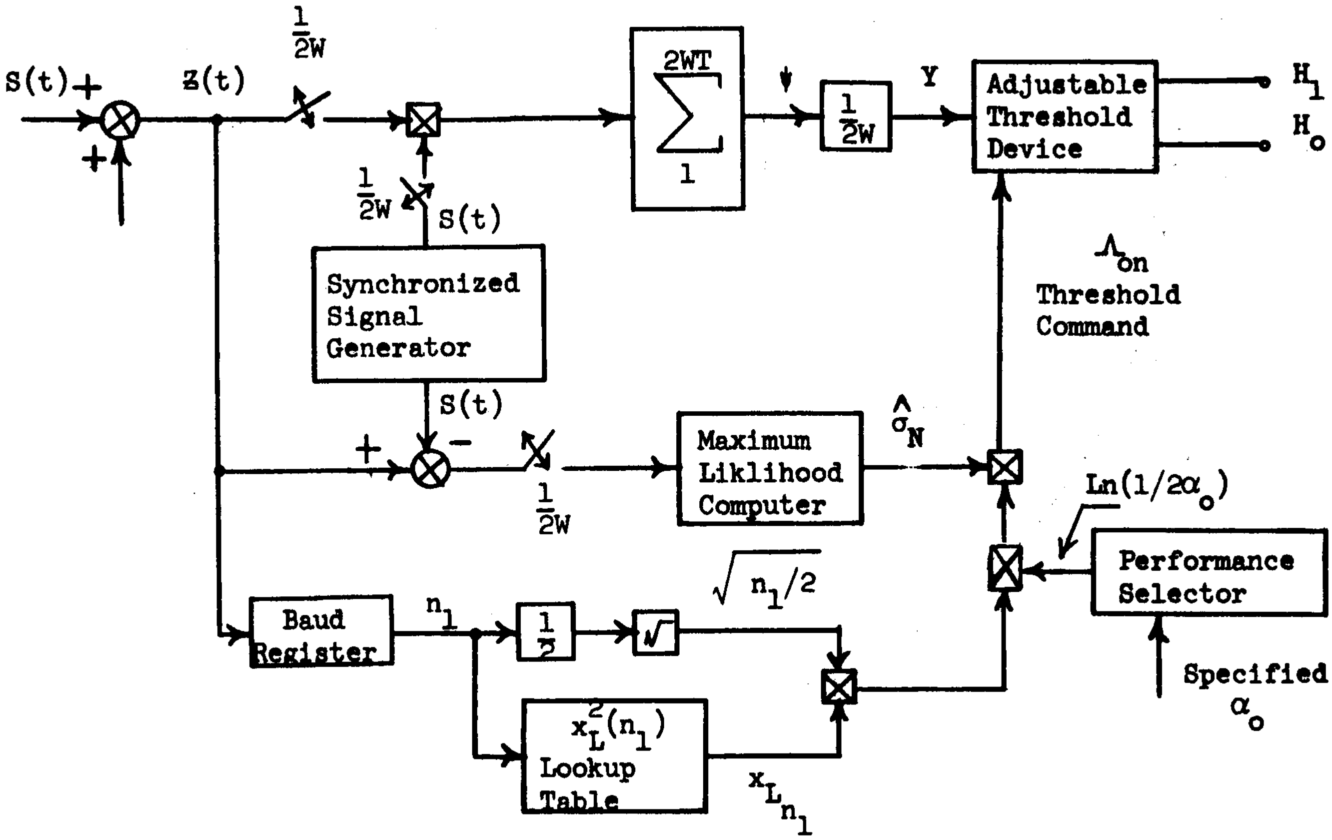

Fig. 9

The error in approximating $P(Y_1)$ with the Laplace density function is defined as

$$\epsilon = | \alpha_n(t) - \alpha_L(t) | \tag{54}$$

where

$$| \alpha_n(t) - \alpha_L(t) | = \left| \int_t^\infty \frac{1}{\sqrt{2\pi}} e^{-v^2} dv - \int_t^\infty \frac{1}{\sqrt{2}} e^{-\sqrt{2}v} dv \right| \tag{55}$$

and is plotted as a function of α_o in Fig. 10.

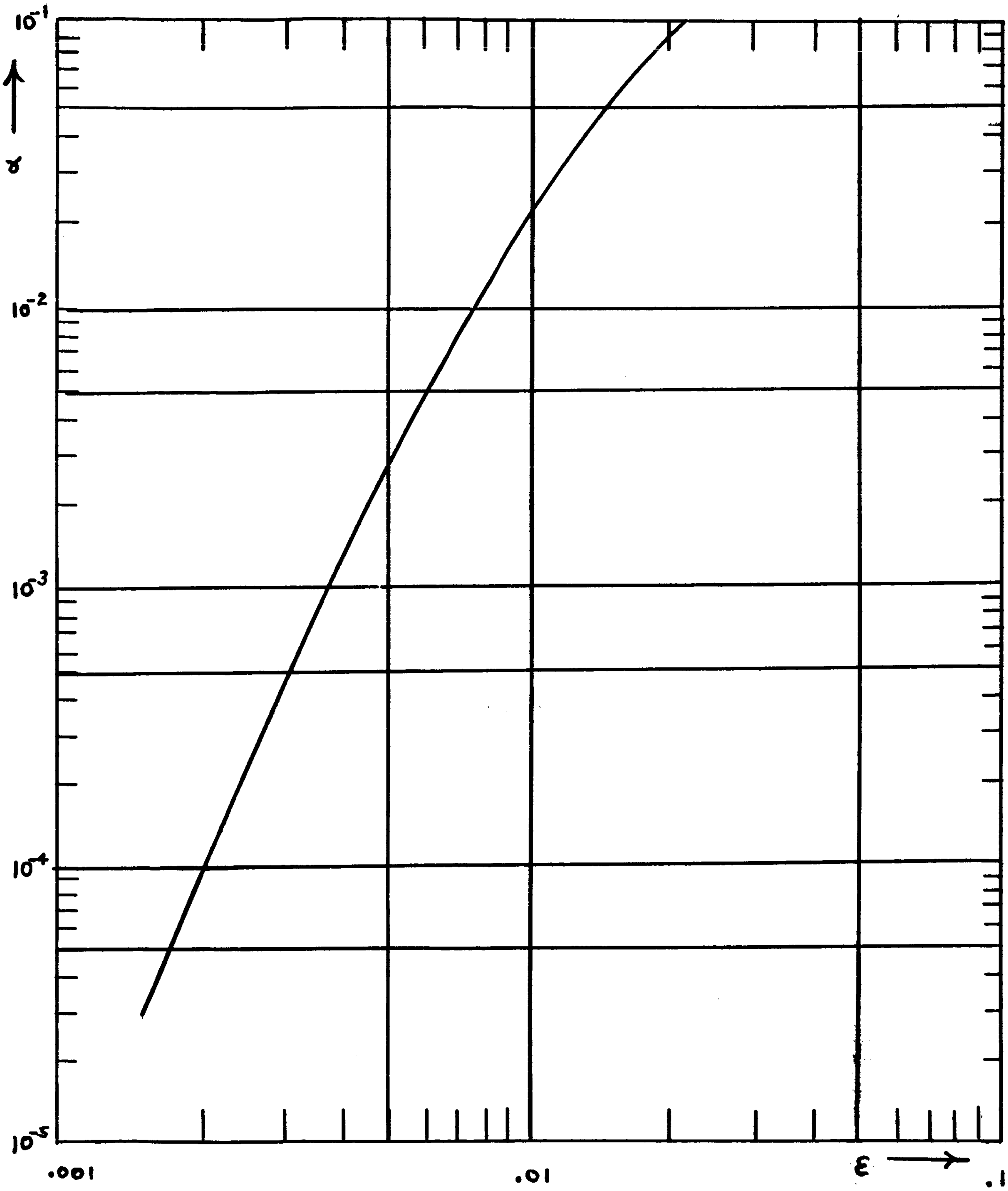

Fig. 10

V. CONCLUSION

For the assumption of a gaussian signal in additive gaussian noise, with unknown covariance, the results of Section III indicate that the critical term in the aposteriori probability expression arrived at through Bayes' theorem is the Wishart density, $W(2WT, n_1, \Phi_1)$, with parameters n_1, Φ_1 as defined. With the additional assumptions of white, bandlimited signal and noise, classified data at the receiver, and the discussion in Section IV, it was shown that even with the Wishart form, the aposteriori probability is maximized when the norm, $\|\underline{Z}-\underline{S}_i\|$, is minimized, leading to the correlation computer structure. The results of Section III were used to illustrate the up dating of Λ, with classified estimates.

Section IV introduced a technique for selecting a threshold, Λ_{o_n}, for hypothesis testing. It was shown that actual detector performance, α, would be a random variable with a mean and variance that converges to the specified performance, α_o, and zero respectively. For the case of apriori knowledge of Λ_{n_1} based upon sample size n_o, Figs. 7 and 8 may be entered at a point $n_o + n_1$ to establish the uncertainty of performance resolved by apriori knowledge. If Λ_o is a guess, no confidence can be assigned, and the convergence properties, as established, will not hold.

The curves in Figs. 7 and 8 are normalized with respect to $\mathrm{Ln}\ 1/2\alpha_o$ to illustrate convergence properties independent of SNR. The results indicate that with no apriori knowledge, 23 supervisory bauds are sufficient to yield expected performance to within 10% of specified performance and 97 bauds yield expected performance to within 2% of specified performance. The normalized variance of performance is down to 25% by the 23rd baud and decreases to 4.5% in 99 bauds. The convergence characteristics of the 99 percentile upper confidence limit for a weighed sample variance and a weighed variance of a sample variance are included for comparison. The rapid convergence of $E[h]/(\ln 1/2\alpha_o)^2$ for small values of n_1 is due to the difference of gamma functions in the expression for $E[h]$.

The error, ϵ in computing actual α due to the approximation ranges from 2.2% for specified $\alpha_o = 10^{-1}$ to .15% for $\alpha_o = 3.X10^{-5}$, indicating the nature of the Laplace density as an approximation in this case.

REFERENCES

[1] H. Cramér, "Mathematical Methods of Statistics," Princeton University Press, Princeton, N. J., pp. 406, 1961

[2] T. W. Anderson, "An Introduction to Multivariate Analysis," John Wiley and Sons, New York, N. Y., pp.67, 154-163, 1958

[3] D. G Keehn, "Learning the Mean and Covariance Matrix of Gaussian Signals in Pattern Recognition," Stanford Electronics Laboratory, Technical Report No. 2003-6, February 1963

[4] C. R. Rao, "Advanced Statistical Methods in Biometric Research," John Wiley and Sons, New York, N. Y., 1952

[5] W. Gröbner and N. Hofreiter, "Integral tafel II Teil," Springer-Verlog, pp. 11, 1961

APPENDIX I

The terms in Bayes' theorem for the gaussian case are obtained by considering the conditional probability that the mth received sample falls in an interval dz_m as

$$P\left\{z_m | s_{i_m}\right\} = P(z_m | s_{i_m}) dz_m \quad \frac{1}{\sqrt{2\pi}\sigma_N} \quad e^{-\frac{\left(z_m - s_{i_m}\right)^2}{2\sigma^2}} dz_m \qquad \text{(I-1)}$$

which for sample size $2WT$ of identically distributed samples

$$P(\underline{z} | \underline{S}_i) = \left[2\pi\right]^{-\frac{2WT}{2}} \left|\Lambda_n\right|^{-\frac{1}{2}} e^{-\frac{1}{2}[\underline{z} - \underline{S}_i]' \Lambda_n^{-1} [\underline{z} - \underline{S}_i]} \prod_{m=1}^{2WT} dz_m \qquad \text{(I-2)}$$

Likewise

$$P\left\{\underline{z}\right\} = P(\underline{z}) \prod_{m=1}^{2WT} dz_m = \sum_{i=1}^{2} P(\underline{z}, \underline{S}_i) \prod_{m=1}^{2WT} dz_m$$

$$= \left[\sum_{i=1}^{2} P(\underline{S}_i) \; P(\underline{z} | \underline{S}_i)\right] \prod_{m=1}^{2WT} dz_m$$

$$= \left[P(\underline{S}_1) \; P(\underline{z} | \underline{S}_1) + P(\underline{S}_2) \; P(\underline{z} | \underline{S}_2)\right] \prod_{m=1}^{2WT} dz_m \qquad \text{(I-3)}$$

and $P\left\{\underline{S}_i | \underline{z}\right\}$ follows as in (4).

The means and variances of the detector computed hypothesis testing variables are found as follows.

The detector branch matched for s(t) will compute variables ψ_1 and ψ_2 depending upon whether signal is absent or present respectively.

Thus

$$E\left[\psi_1\right] = E\left[\sum_{m=1}^{2WT} z_m s_m\right] = E\left[\sum_{m=1}^{2WT} n_m s_m\right] = \sum_{m=1}^{2WT} E\left[n_m s_m\right]$$

$$= \sum_{m=1}^{2WT} E\left[n_m\right] E\left[s_m\right] = 0 \qquad \text{(I-4)}$$

for identically distributed, 2WT variate, zero mean samples where signal and noise are independent.

$$E\left[\psi_1^2\right] = E\left[(\underline{Z}'\underline{S})^2\right] = E\left[(\underline{N}'\underline{S})^2\right] = E\left[\sum_{m=1}^{2WT} n_m s_m \sum_{k=1}^{2WT} n_k s_k\right]$$

$$= E\left[\sum_{m=1}^{2WT} (n_m s_m)^2 + \sum_{m=1}^{2WT} \sum_{\substack{k=1 \\ k\neq m}}^{2WT} n_m s_m n_k s_k\right]$$

$$= \sum_{m=1}^{2WT} E\left[n_m^2 s_m^2\right] + \sum_{m=1}^{2WT} \sum_{\substack{k=1 \\ k\neq m}}^{2WT} E\left[n_m s_m n_k s_k\right]$$

$$= \sum_{m=1}^{2WT} E\left[n_m^2\right] E\left[s_m^2\right] + \sum_{m=1}^{2WT} \sum_{\substack{k=1 \\ k\neq m}}^{2WT} E\left[n_m\right] E\left[s_m\right] E\left[n_k\right] E\left[s_k\right]$$

$$= 2WT\ \sigma_N^2\ \sigma_s^2 + 0 \qquad \text{(I-5)}$$

where samples m, k are uncorrelated and hence independent.

$$E\left[\psi_2\right] = E\left[\sum_{m=1}^{2WT} z_m s_m\right] = E\left[\sum_{m=1}^{2WT} s_m^2 + n_m s_m\right] = \sum_{m=1}^{2WT} E\left[s_m^2\right] + E\left[n_m\right] E\left[s_m\right]$$

$$= 2WT\ \sigma_s^2 = 2WT\ S_o W \qquad \text{WATTS} \qquad \text{(I-6)}$$

for the zero mean assumption.

$$E\left[\psi_2^2\right] = E\left[(\underline{Z}'\underline{S})^2\right] = E\left[[(\underline{S} + \underline{N})'\underline{S}]^2\right] = E\left[(\underline{S}'\underline{S} + \underline{N}'\underline{S})^2\right]$$

$$= E\left[(\underline{S}'\underline{S})^2 + 2(\underline{S}'\underline{S})(\underline{N}'\underline{S}) + (\underline{N}'\underline{S})^2\right]$$

$$= E\left[(\underline{S}'\underline{S})^2\right] + 2\,E\left[(\underline{S}'\underline{S})(\underline{N}'\underline{S})\right] + E\left[(\underline{N}'\underline{S})^2\right] \qquad \text{(I-7)}$$

Now

$$E\left[(\underline{S}'\underline{S})^2\right] = E\left[\sum_{m=1}^{2WT} s_m^2 \sum_{k=1}^{2WT} s_k^2\right] = E\left[\sum_{m=1}^{2WT} s_m^4 + \sum_{m=1}^{2WT} \sum_{\substack{k=1 \\ k\neq m}}^{2WT} s_m^2\ s_k^2\right]$$

$$= \sum^{2WT} E\left[s_m^4\right] + \sum_{m=1}^{2WT} \sum_{\substack{k=1 \\ k \neq m}}^{2WT} E\left[s_m^2\right] E\left[s_k^2\right]$$

$$= 2WT\ \mu_4' + 2WT(2WT-1)\ \sigma_s^2\ \sigma_s^2 \tag{I-8}$$

For the zero mean gaussian assumptions, the rth raw moment is

$$\mu_r' = 1.3.5\ldots(r-1)\ \sigma^r, \text{ hence}$$
$$\mu_4' = 3\sigma^4 \tag{I-9}$$

and

$$E\left[(\underline{S}'\underline{S})^2\right] = 2WT\ (3\sigma_s^4) + 2WT\ (2WT-1)\ \sigma_s^4 \tag{I-10}$$

$$E\left[(\underline{S}'\underline{S})(\underline{N}'\underline{S})\right] = E\left[\sum_{m=1}^{2WT} s_m^2 \sum_{k=1}^{2WT} n_k s_k\right] = E\left[\sum_{m=1}^{2WT} s_m^3\ n_m + \sum_{m=1}^{2WT} \sum_{\substack{k=1 \\ m \neq k}}^{2WT} s_m^2\ n_k s_k\right]$$

$$= \sum_{m=1}^{2WT} E\left[s_m^3\ n_m\right] + \sum_{m=1}^{2WT} \sum_{\substack{k=1 \\ m \neq k}}^{2WT} E\left[s_m^2\right] E\left[n_k\right] E\left[s_k\right]$$

$$= \sum_{m=1}^{2WT} E\left[s_m^3\right] E\left[n_m\right] + 0$$

$$= 0 \tag{I-11}$$

From (I-5), $E\left[(\underline{N}'\underline{S})^2\right] = 2WT\ \sigma_s^2\ \sigma_N^2$ (I-12)

Hence from (I-7), (I-10), (I-11) and (I-12),

$$E\left[\psi_2^2\right] = 2WT\ (3\sigma_s^4) + 2WT\ (2WT-1)\ \sigma_s^4 + 2WT\ \sigma_s^2\ \sigma_N^2 \tag{I-13}$$

The variances are

$$\sigma_{\psi_1}^2 = E\left[\psi_1^2\right] - E^2\left[\psi_1\right] = 2WT\ \sigma_N^2\ \sigma_s^2 - 0 \tag{I-14}$$

$$\sigma_{\psi_2}^2 = E\left[\psi_2^2\right] - E^2\left[\psi_2\right] =$$

$$= 2WT \left[3\sigma_s^4 + (2WT-1)\,\sigma_s^4 + \sigma_s^2\,\sigma_N^2 \right] -2WT\,(2WT)\,\sigma_s^4$$

$$= 2WT \left[3\sigma_s^4 + 2WT\,\sigma_s^4 - \sigma_s^4 + \sigma_s^2\,\sigma_N^2 - 2WT\,\sigma_s^4 \right]$$

$$= 2WT \left[\sigma_s^2\,(2\sigma_s^2 + \sigma_N^2) \right] \qquad \text{(I-15)}$$

APPENDIX II

Using Baye's theorem, the aposteriori, probability for the binary case is

$$P\left\{\underline{S}_i \mid \underline{Z}, \underline{z}\right\} = \frac{P\left\{\underline{S}_i\right\} P(\underline{Z}, \underline{z} \mid \underline{S}_i)}{\sum_{i=1}^{2} P\left\{\underline{S}_i\right\} P(\underline{Z}, \underline{z} \mid \underline{S}_i)} \tag{II-1}$$

which is maximized when $P(\underline{Z}, \underline{z} \mid \underline{S}_i)$ is maximum.

But

$$P(\underline{Z}, \underline{z} \mid \underline{S}_i) = P(\underline{z} \mid \underline{Z}; \underline{S}_i)\, P(\underline{Z} \mid \underline{S}_i) \tag{II-2}$$

and

$$P(\underline{z} \mid \underline{Z}; \underline{S}_i) = \int P(\underline{z} \mid \underline{Z}; \theta, \underline{S}_i)\, P(\theta \mid \underline{Z}; \underline{S}_i)\, d\theta \tag{II-3}$$

where θ is the inverse of the covariance matrix of the random process $\underline{Z}(t)$ when the ith signal is present. The term $P(\underline{z} \mid \underline{Z}; \theta, \underline{S}_i)$ is the same as $P(\underline{z} \mid \theta; \underline{S}_i)$ and

$$P(\underline{z} \mid \theta; \underline{S}_i) = (2\pi)^{-\frac{k}{2}} \mid \theta \mid^{\frac{1}{2}} e^{-\frac{1}{2} [\underline{z}-\underline{S}_i]^1 \theta [\underline{z}-\underline{S}_i]} \tag{II-4}$$

The remaining team in the integrand is

$$P(\theta \mid \underline{Z}; \underline{S}_i) = \frac{P(\theta)\, P(\underline{Z} \mid \theta; \underline{S}_i)}{\int P(\underline{Z} \mid \theta; \underline{S}_i) P(\theta) d\theta} \tag{II-5}$$

and

$$\underline{z}_j = \begin{bmatrix} z_1 \\ \vdots \\ z_k \end{bmatrix}_j , \quad \underline{Z} = \left\{\underline{z}_1\ \underline{z}_2\ \cdots\ \underline{z}_n\right\} \tag{II-6}$$

$P(\theta)$ is found via the wishart law, which in this case is the n variate analog of taking $f(\hat{\sigma}^2)$ written as

$$f(\hat{\sigma}^2; 1/\sigma 2) = \frac{\left(\frac{k-1}{2}\right)^{\frac{k-1}{2}}}{\Gamma\left(\frac{k-1}{2}\right)} \left(\frac{1}{\sigma^2}\right)^{\frac{k-1}{2}} e^{-\frac{k-1}{2} \hat{\sigma}^2 \left(\frac{1}{\sigma^2}\right)} \qquad 0 < \hat{\sigma}^2 < \infty \tag{II-7}$$

where $$\hat{\sigma}^2 = \sum_{j=1}^{k} \frac{[(Z-S_1)_j - \overline{(Z-S_1)_j}]^2}{k-1}$$

and normalizing by introducing a term c to insure

$$\int_0^\infty cf(\hat{\sigma}^2;1/\sigma^2)\, dp = 1 \tag{II-8}$$

where $p = 1/\sigma^2$

thus

$$P(\theta) = C_{k,n_o} \left|\frac{n_o}{2} \Phi_o\right|^{\frac{n_o-1}{2}} |\theta|^{\frac{n_o-k-2}{2}} e^{-\frac{1}{2} \mathrm{TR}\, n_o \Phi_o \theta} \tag{II-9}$$

where C_{k,n_o} is the required normalizing term developed by Wishart

$$C_{k,n_o} = \frac{1}{\pi^{\frac{k(k-1)}{4}} \prod_{q=1}^{k} \Gamma\left(\frac{n_o-q}{2}\right)} \tag{II-10}$$

and Φ_o and θ are the multivariate analogs of $\hat{\sigma}^2$ and $1/\sigma^2$ respectively. Φ_o is an estimate of θ based upon n_o initial samples from a population containing θ with confidence consistent with n_o <u>or</u> Φ_o is simply an initial guess weighted by n_o.

Using (II-4)

$$P(\underline{Z} \mid \theta;\underline{S}_1) = \prod_{j=1}^{n} P(\underline{z}_j \mid \theta;\underline{S}_1) = \prod_{j=1}^{n} (2\pi)^{\frac{-k}{2}} |\theta|^{1/2} e^{-\frac{1}{2}[\underline{z}-\underline{S}_1]_j' \theta [\underline{z}-\underline{S}_1]_j}$$

$$= (2\pi)^{\frac{-nk}{2}} |\theta|^{\frac{n}{2}} e^{-\frac{1}{2} \mathrm{TR}\, n\, \theta < [\underline{z}-\underline{S}_1][\underline{z}-\underline{S}_1]' >} \tag{II-11}$$

for the $\underline{z}_j$ identically distributed, with

$$< [\underline{z}-\underline{S}_1][\underline{z}-\underline{S}_1]' > = \frac{1}{n} \sum_{j=1}^{n} (\underline{z}-\underline{S}_1)_j (\underline{z}-\underline{S}_1)'_j \tag{II-12}$$

Neglecting the denominator of (II-5) since it only serves to normalize $P(\theta \mid \underline{Z};\underline{S}_1)$ to satisfy $\int d\theta = 1$ and using (II-9), (II-11),

$$P(\theta|\underline{Z};\underline{S}_1) \approx P(\theta)\ P(\underline{Z}|\theta;\underline{S}_1)$$

$$= C_{k,n_o}\left|\frac{n_o}{2}\Phi_o\right|^{\frac{n_o-1}{2}}|\theta|^{\frac{n_o-k-2}{2}} e^{-\frac{1}{2}\mathrm{TR}\ n_o\Phi_o\theta}\ )(\ (2\pi)^{-\frac{nk}{2}}|\theta|^{\frac{n}{2}} e^{-\frac{1}{2}\mathrm{TR}\ n\theta <[\underline{Z}-\underline{S}_1][\underline{Z}-\underline{S}_1]'>}$$

$$= C_{k,n_o}\left|\frac{n_o}{2}\Phi_o\right|^{\frac{n_o-1}{2}}|\theta|^{\frac{n_o+n-k-2}{2}}(2\pi)^{-\frac{nk}{2}} e^{-\frac{1}{2}(\mathrm{TR}\ n_o\Phi_o\theta+\mathrm{TR}n <[\underline{Z}-\underline{S}_1][\underline{Z}-\underline{S}_1]'>)}$$

$$= C_{k,n_o}(2\pi)^{-\frac{nk}{2}}\left|\frac{n_o}{2}\Phi_o\right|^{\frac{n_o-1}{2}}|\theta|^{\frac{n_o+n-k-2}{2}} e^{-\frac{1}{2}\mathrm{TR}\ \theta(n_o\Phi_o+n <[\underline{Z}-\underline{S}_1][\underline{Z}-\underline{S}_1]'>)}$$

$$= C_{k,n_o}(2\pi)^{-\frac{nk}{2}}\left|\frac{n_o}{2}\Phi_o\right|^{\frac{n_o-1}{2}}|\theta|^{\frac{n_o+n-k-2}{2}} e^{-\frac{1}{2}\mathrm{TR}(n_o+n)\theta\frac{(n_o\Phi_o+n <[\underline{Z}-\underline{S}_1][\underline{Z}-\underline{S}_1]'>)}{(n_o+n)}} \tag{II-13}$$

Thus $P(\theta|\underline{Z};\underline{S}_1)$ is a Wishart density, $W(k,n_1,\Phi_1)$ with parameters

$$n_1 = n_o+n$$

$$\Phi_1 = \frac{n_o\Phi_o+n < [\underline{Z}-\underline{S}][\underline{Z}-\underline{S}_1]' >}{n_o+n} \tag{II-14}$$

Substituting n_1, Φ_1 into (II-13) and modifying C_{k,n_o} to absorb the effects and retain normality

$$P(\theta|\underline{Z};\underline{S}_1) = C_{k,n_1}\left|\frac{n_1}{2}\Phi_1\right|^{\frac{n_1-1}{2}}|\theta|^{\frac{n_1-k-2}{2}} e^{-\frac{1}{2}\mathrm{TR}\ n_1\Phi_1\theta} \tag{II-15}$$

Using (II-4) and (II-15), the <u>integral</u> of (II-3) is

$$P(\underline{Z}|\theta;\underline{S}_1)\ P(\theta|\underline{Z};\underline{S}_1) =$$

$$(2\pi)^{-\frac{k}{2}}|\theta|^{1/2} e^{-\frac{1}{2}\mathrm{TR}\ \theta[\underline{Z}-\underline{S}_1][\underline{Z}-\underline{S}_1]'} \cdot C_{k,n_1}\left|\frac{n_1}{2}\Phi_1\right|^{\frac{n_1-1}{2}}|\theta|^{\frac{n_1-k-2}{2}}\ )(\ e^{-\frac{1}{2}\mathrm{TR}\ n_1\Phi_1\theta}$$

$$= (2\pi)^{-\frac{k}{2}} C_{k,n_1}\left|\frac{n_1}{2}\Phi_1\right|^{\frac{n_1-1}{2}} e^{-\frac{1}{2}\mathrm{TR}[(\underline{Z}-\underline{S}_1)(\underline{Z}-\underline{S}_1)'+n_1\Phi_1]\theta}|\theta|^{\frac{n_1-k-2}{2}} \tag{II-16}$$

Using (II-16) into (II-3)

$$P(\underline{Z} \mid \underline{Z};\underline{S}_1) = (2\pi)^{\frac{-k}{2}} C_{k,n_1} \left| \frac{n_1}{2} \Phi_1 \right|^{\frac{n_1-1}{2}} \int \cdots \int |\theta|^{\frac{n_1-k-2}{2}} e^{-\frac{1}{2} TR[(\underline{Z}-\underline{S}_1)'(\underline{Z}-\underline{S}_1) + n_1\Phi_1]\,\theta} d\theta_{11} \cdots d\theta_{kk} \quad \text{(II-17)}$$

But an identity formulated by Cramer has

$$\int \cdots \int |\theta|^{\frac{n_1-k-1}{2}} e^{-TRA\theta} d\theta_{11} \cdots d\theta_{kk} = \frac{1}{C_{k,n_1} |A|^{\frac{n_1}{2}}} \quad \text{(II-18)}$$

Thus

$$P(\underline{Z} \mid \underline{Z};\underline{S}_1) = (2\pi)^{\frac{-k}{2}} C_{k,n_1} \left| \frac{n_1}{2} \Phi_1 \right|^{\frac{n_1-1}{2}} \frac{1}{C_{k,n_1} \left| \frac{1}{2} ([\underline{Z}-\underline{S}_1][\underline{Z}-\underline{S}_1]' + n_1\Phi_1) \right|^{\frac{n_1}{2}}}$$

$$= \frac{(2\pi)^{\frac{-k}{2}} \left| \frac{n_1}{2} \Phi_1 \right|^{\frac{n_1-1}{2}}}{\left| \frac{1}{2} ([\underline{Z}-\underline{S}_1][\underline{Z}-\underline{S}_1]' + n_1\Phi_1) \right|^{\frac{n_1}{2}}} \quad \text{(II-19)}$$

Assuming Φ_1 is non singular, Φ_1^{-1} exists, and hence

$$\frac{1}{2} ([\underline{Z}-\underline{S}_1][\underline{Z}-\underline{S}_1]' + n_1\Phi_1) = \left([\underline{Z}-\underline{S}_1][\underline{Z}-\underline{S}_1]' \frac{\Phi_1^{-1}}{n_1} + I\right) \frac{n_1}{2} \Phi_1 \quad \text{(II-20)}$$

Factoring the numerator of (II-19) and substituting (II-20) yields

$$P(\underline{Z} \mid \underline{Z};\underline{S}_1) = \frac{(2\pi)^{\frac{-k}{2}} \left| \frac{n_1}{2} \Phi_1 \right|^{-\frac{1}{2}} \left| \frac{n_1}{2} \Phi_1 \right|^{\frac{n_1}{2}}}{\left| [\underline{Z}-\underline{S}_1][\underline{Z}-\underline{S}_1]' \frac{\Phi_1^{-1}}{n_1} + I \right|^{\frac{n_1}{2}} \left| \frac{n_1}{2} \Phi_1 \right|^{\frac{n_1}{2}}}$$

$$= \frac{(2\pi)^{\frac{-k}{2}} \left| \frac{n_1}{2} \Phi_1 \right|^{-\frac{1}{2}}}{\left| [\underline{Z}-\underline{S}_1][\underline{Z}-\underline{S}_1]' \frac{\Phi_1^{-1}}{n_1} + I \right|^{\frac{n_1}{2}}} \quad \text{(II-21)}$$

For the conditions of the derivation, the identities (II-22)

$$\left| \frac{[\underline{z}-\underline{S}_i][\underline{z}-\underline{S}_i]'\Phi_1^{-1}}{n_1} + I\right|^{\frac{n_1}{2}} = \left(\frac{1}{n_1} \, TR(\underline{z}-\underline{S}_i)(\underline{z}-\underline{S}_i)' \, \Phi_1^{-1} + 1\right)^{+\frac{n_1}{2}}$$

Holds, thus

$$P(\underline{z}|\underline{Z};\underline{S}_i) = \frac{(2\ )^{-\frac{k}{2}}}{\left(\frac{n_1}{2}\right)^{1/2}} \left|\theta_1^{-1}\right|^{-\frac{1}{2}} \left(1 + \frac{1}{n_1} \, TR(\underline{z}-\underline{S}_i)\,(\underline{z}-\underline{S}_i)' \, \Phi_1^{-1}\right)^{-\frac{n_1}{2}} \qquad (II-23)$$

and when normalized, becomes

$$P(\underline{z}|\underline{Z};\underline{S}_i) = \frac{\Gamma\left(\frac{n_1}{2}\right)}{\Gamma\left(\frac{n_1-k}{2}\right)} \frac{(2\pi)^{-\frac{k}{2}}}{\left(\frac{n_1}{2}\right)^{1/2}} \left|\Phi_1^{-1}\right|^{-1/2} \left(1 + \frac{1}{n_1} \, TR(\underline{z}-\underline{S}_i)(\underline{z}-\underline{S}_i)'\Phi_1^{-1}\right)^{-\frac{n_1}{2}} \qquad (II-24)$$

Thus (II-1) is maximized when (II-24) is maximized, which occurs for minimum values of $TR(\underline{z}-\underline{S}_i)(\underline{z}-\underline{S}_i)' \, \Phi_1^{-1}$

But

$$TR(\underline{z}-\underline{S}_i)(z-s_i)'\Phi_1^{-1} = (\underline{z}-\underline{S}_i)'\Phi_1^{-1} \, (\underline{z}-\underline{S}_i) \qquad (II-25)$$